COS 300045 – DATA VISUALISATION  
  
Project Stand Up 1  
  
DESIGN BOOK

NAME – Yulan Yaneth  
ID – 105107919

Table of Contents

1. Introduction.........................................................................2

1.1 Background................................................................2

1.2 Motivation..................................................................2

1.3 Purpose.......................................................................3

2. Exploratory Data Analysis...................................................4

2.1 Study Time vs Final Grade.......................................4

2.2 Past Failures vs Final Grade....................................5

2.3 Absenteeism vs Final Grade..................................6

3. Variable Types & Data Preparation...................................7

4. Challenges & Next Steps..................................................8

5. KNIME Workflow Snapshot..............................................9

# 1. Introduction

## 1.1 Background

Academic aptitude, social context, familial support, and behavioral tendencies all influence student success, which is a multifaceted result. Finding students who are likely to do poorly or drop out for unknown reasons is a common problem for educational institutions. Conventional measures, like final grades or attendance records, frequently don't offer early warning signs, which makes prompt treatments challenging. In response, the goal of this project is to use data visualization to find trends in student performance and empower teachers to help at-risk pupils in a proactive manner.

## 1.2 Motivation

Many underachievers don't lack potential; they're just overlooked until it's too late. Instructors might not have the time or resources to thoroughly analyze student datasets. By offering an easy-to-use interface for examining and highlighting early warning indicators like absenteeism, low engagement, and recurrent failures, visual analytics help close this gap. By transforming raw data into understandable visual narratives that promote student success and welfare, this initiative aims to increase the visibility and actionability of performance indicators.

## 1.3 Purpose

The purpose of this project is to design an interactive visualisation tool that assists educators in analysing student performance data. Using a dataset from the UCI Machine Learning Repository, the tool focuses on identifying at-risk students based on various academic and behavioural factors. Key questions the dashboard addresses include:  
- Does increased study time correlate with higher grades?  
- Are students with multiple past failures likely to continue struggling?  
- How significantly do absences affect final performance?  
By answering these questions, the visualisation serves as a decision-making tool for teachers, school counsellors, and academic policy advisors.

# 2. Exploratory Data Analysis

## 2.1 Study Time vs Final Grade

The relationship between the time students dedicate to studying and their academic performance is a critical indicator of learning behaviour. Using KNIME, the dataset was grouped by the attribute `studytime`, and the average of the final grade (`G3`) was calculated for each group. The bar chart clearly showed a positive trend: students who studied less (group 1) scored an average of 10.0, whereas those in the highest study group scored above 11.2. This insight supports the widely held view that study habits are a key predictor of academic success.

## 2.2 Past Failures vs Final Grade

To assess whether historical academic challenges continue to impact student outcomes, the dataset was analysed by grouping the `failures` attribute. Students who had never failed previously averaged 11.25 in their final grades. In contrast, those with three past failures had an average of only 5.69. This steep decline illustrates that recurring academic struggles are often not isolated and require early academic interventions to prevent long-term decline.

## 2.3 Absenteeism vs Final Grade

The third insight assessed how student absences affected academic achievement. Four groups of absences were identified: 0–2, 3–5, 6–10, and 11+. While students with 11+ absences averaged only 8.5, those with fewer absences (0–2) received an average grade of 12.3. The unmistakable downward trend demonstrates how important regular attendance in class is to sustaining academic success. Thus, absenteeism can serve as a warning sign for disengagement or outside difficulties.

# 3. Variable Types & Data Preparation

The dataset consists of different data types that influence how analysis is performed:  
- Categorical: gender, address, school support (schoolsup), family support (famsup), internet access, activities, and guardian  
- Ordinal: studytime, failures, education level of parents  
- Ratio: G1, G2, G3 (grades), absences, age  
  
During data preparation in KNIME, transformations included:  
- Converting string types to categorical for group-by aggregation  
- Binning continuous variables like absences into logical ranges using the Rule Engine  
- Ensuring numerical columns were correctly formatted for statistics and bar charts  
This step ensured compatibility with visualisation nodes and enabled meaningful summaries.

# 4. Challenges & Next Steps

Challenges encountered during the early stages included:  
- CSV format inconsistencies that affected import into KNIME  
- Determining appropriate binning methods for numeric columns (e.g., absences)  
- Selecting visualisations that are both meaningful and simple to interpret  
  
Next steps involve:  
- Adding a fourth insight (e.g., family support vs grade)  
- Refining dashboard interactivity  
- Completing usability feedback and submitting final report via GitHub

# 5. KNIME Workflow Snapshot

Below is a visual overview of the KNIME workflow used to conduct data aggregation, transformation, and visualisation:

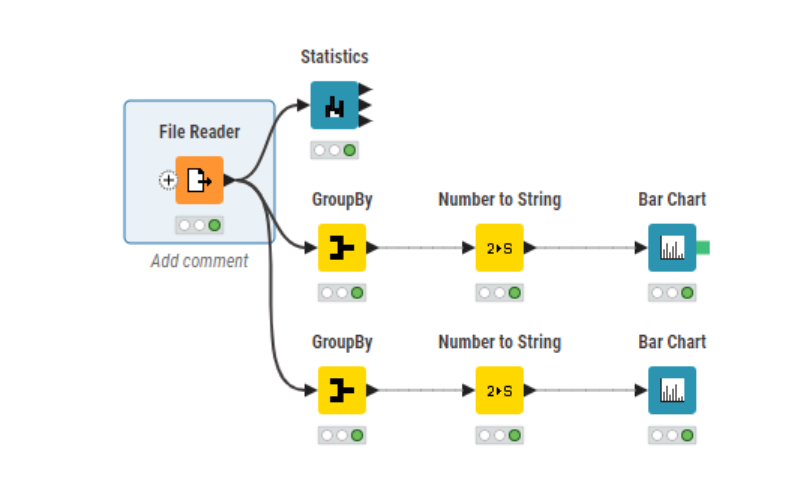


Figure 1: KNIME workflow with GroupBy, conversion, and Bar Chart nodes.